Nuevos ejercicios propuestos

1. El calcio se encuentra presente en la sangre de los mamíferos sanos en concentraciones que siguen una distribución normal de media 6 mg. por cada 100 ml. de sangre, y una desviación típica de 1 mg. por cada 100 ml. de sangre. Un individuo con una variabilidad mayor podría tener graves trastornos de coagulación, por lo que interesa controlarla. Una serie de 9 pruebas con el mismo individuo dieron una media y una desviación muestrales de 6.2 y 2 mg. de calcio por cada 100 ml. de sangre, respectivamente. Se pregunta:

   (a) ¿Se puede considerar que la desviación típica de la cantidad de calcio del individuo es significativamente superior al miligramo por cada 100 ml. de sangre? (utilizar $\alpha = 0.05$ como nivel de significación).

   (b) ¿Hay evidencia, usando $\alpha = 0.05$ como nivel de significación, de que la cantidad media de calcio de este individuo sea diferente de la habitual en los mamíferos sanos?

2. Se desea realizar un estudio sobre la permeabilidad de cierto terreno, que se supone que se distribuye según una normal de media μ y desviación conocida $\sigma = 2.5$ (en mm. por segundo). Estudios previos poco fiables indicaban que el valor de μ era 80 mm. por segundo, pero se sospecha que en realidad esto no es así. Con la finalidad de dar solidez a esta sospecha, se hace lo siguiente:

   (a) Se fija un valor mayor que 80, por ejemplo 82.5. Si realmente μ fuese 80, se pide calcular la probabilidad de que la permeabilidad de una extracción de terreno tomada al azar sea mayor que 82.5.

   (b) Se toma una muestra formada por 200 extracciones del terreno hechas al azar. Para ahorrar tiempo y dinero, se hace lo siguiente:

para cada extracción sólo se observa si su permeabilidad es o no superior a 82.5 mm. por segundo (si pasado un segundo no ha llegado a 82.5 mm., es inferior, y si antes del segundo llega a 82.5, podemos dejar de observar porque entonces será superior). Resulta que de las 200 extracciones analizadas, sólo hay 53 para las que su permeabilidad es superior a 82.5 mm. por segundo. Se pide calcular el intervalo de confianza asintótico con nivel de confianza (aproximado) de $\gamma = 0.95$ para la proporción de extracciones de terreno cuya permeabilidad es superior a los 82.5 mm. por segundo.

(c) Comparando la probabilidad encontrada en el apartado a) con el intervalo de confianza del apartado b), ¿se puede pensar que la sospecha queda reafirmada? (justificar la respuesta).

3. Algunas compañías aéreas han calculado en un 14 % la proporción de personas de entre las que reservan un billete de avión que después no se presentan o cambian de vuelo. En consecuencia, para no perder dinero, estas compañías venden más plazas de las realmente disponibles (*overbooking*). Se pide:

   (a) Si una compañía ha aceptado 240 reservas para un vuelo en el que sólo hay 213 plazas disponibles, ¿cuál es, aproximadamente, la probabilidad de disponer finalmente de una plaza para todos los pasajeros que tienen una reserva y se presentan al vuelo?

   (b) Si las personas que viajan en avión pierden un 2 % de los vuelos que deben tomar por causa de *overbooking* y una persona realiza 20 vuelos en un año, ¿qué probabilidad tiene de perder dos o más vuelos en ese año?

4. La siguiente tabla muestra los datos recogidos en una investigación sobre el consumo de marihuana entre los estudiantes de secundaria de cierta zona. En ella se clasificaron 445 estudiantes de secundaria según su consumo de esta sustancia (Nunca, Ocasional o Regular), y también según el consumo habitual de bebidas alcólicas y/o de drogas por parte de los padres (Ninguna, Una de las dos o Ambas cosas):

		Estudiante:		
		Nunca	Ocasional	Regular
	Ninguna	141	54	40
Padres:	Una	68	44	51
	Ambas	17	11	19

A partir de los datos de la tabla se pide:

(a) ¿Podemos decir, con una probabilidad de equivocarnos al hacerlo de un 1 %, que hay relación entre el consumo de marihuana de los estudiantes de secundaria y el consumo habitual de bebidas alcólicas y/o drogas por parte de sus padres?

(b) En caso de que la respuesta del apartado anterior sea afirmativa, ¿en qué sentido es esta relación?

5. Se sabe que el 83 % de los conductores que circulan por una ciudad utilizan el cinturón de seguridad. Se pide:

   (a) Si la policía detiene aleatoriamente a 10 conductores de esta ciudad, ¿cuál es la probabilidad de que ponga más de una multa por no llevar el cinturón de seguridad?

   (b) Si detiene a 200 conductores, ¿cuál es la probabilidad (aproximada) de que ponga más de 40 multas por no llevar el cinturón?

6. Se ha determinado el contenido de grasas saturadas (en porcentaje) en dos muestras independientes de croquetas de queso de dos marcas diferentes. Los resultados obtenidos se resumen en la siguiente tabla:

	Marca A	Marca B
Tamaño muestral	$n_1 = 10$	$n_2 = 16$
Desviación muestral	$s_1 = 1.2$	$s_2 = 2.2$

   Supongamos que el contenido en grasas saturadas en las croquetas sigue una distribución normal. Se pide hacer un test de hipótesis con nivel de significación $\alpha = 0.05$ para decidir si se puede decir que hay más variabilidad en el porcentaje de grasa en la Marca B que en la Marca A. Repetir el test con $\alpha = 0.01$.

7. El ayuntamiento de un municipio desea llevar a cabo un estudio para saber hasta qué punto sus habitantes son conscientes de lo que la recogida selectiva de la basura comporta desde un punto de vista medio-ambiental. Para ello prepara una encuesta que le permita estimar el porcentaje de habitantes (de entre 20 y 50 años de edad) que son conscientes. Se pide:

   (a) Si el ayuntamiento desea estimar este porcentaje con un error no superior al 5 % y una confianza (aproximada) no inferior al 95 %, ¿a cuántos habitantes del pueblo debería realizar la encuesta como mínimo?

   (b) Responder de nuevo a la pregunta anterior pero ahora utilizando la información de la que dispone el ayuntamiento de que el porcentaje es de almenos un 75 %.

   (c) Si finalmente el ayuntamiento decide tomar una muestra de 300 habitantes de entre 20 y 50 años de edad, y resulta que de ellos hay 264 exactamente que se puede considerar que son conscientes, ¿qué estimación (puntual) obtenemos para el porcentaje de habitantes en esta franja de edad que son conscientes? Encontrar también un intervalo de confianza (asintótico) con nivel de confianza $\gamma = 0.95$ (aproximada) para este porcentaje. En vista de este intervalo, ¿resulta creíble la información de la que dispone el ayuntamiento mencionada en el apartado anterior?

8. Diversos trabajos han mostrado que los líquenes son excelentes indicadores biológicos de la contaminación del aire. En un artículo de investigación del año 1993 se muestra un estudio sobre el *Epiphytic Lichen Hypogymnia Physodes* como "biomonitor" del nitrógeno atmosférico y de las deposiciones de sulfuro en Noruega. La información contenida en la tabla son las lecturas de las variables $X =$"cantidad de líquen" (en porcentaje de peso en seco) e $Y =$"cantidad de nitrógeno atmosférico" (en gr/m^2) para 10 zonas diferentes de Noruega consideradas en el estudio:

X	0.45	0.47	0.58	0.69	0.81	0.86	1.00	0.98	1.24	1.42
Y	0.10	0.12	0.31	0.37	0.42	0.58	0.68	0.73	0.85	0.92

   Se pide:

(a) ¿Se puede decir que hay relación lineal significativa entre las dos variables? (Calcular el coeficiente de correlación de Pearson).

(b) Si la respuesta a la pregunta anterior es afirmativa, encontrar los coeficientes de la recta de regresión de Y sobre X.

(c) Dar la predicción a partir de los datos, si se puede, para la cantidad de nitrógeno atmosférico (en gr/m^2) en una zona de Noruega si la cantidad de líquen (en porcentaje de peso en seco) presente es de 0.94.

9. Para estudiar la aceptación entre los consumidores de una variedad transgénica de determinado cereal se ha encuestado a 200 personas que viven en ámbitos urbanos y a 200 que viven en ámbitos rurales, sin relación entre sí. De entre los encuestados, 107 del primer grupo y 134 del segundo estaban a favor de la variedad transgénica. Se pide:

   (a) ¿Podemos decir que las proporciones de consumidores favorables a la variedad transgénica en ámbitos urbanos y en ámbitos rurales, son significativamente diferentes, con una probabilidad de equivocarnos al decirlo de (aproximadamente) un 10 % ?

   (b) ¿Hasta que valor podemos disminuir el nivel de significación α para que a partir de estos datos se pueda aceptar que hay una diferencia significativa entre las dos proporciones?

10. En una planta de reciclaje de vidrio llegan los envases de vidrio para reciclar desde tres puntos diferentes de recogida, digamos A, B y C, en porcentajes respectivos del 35 %, el 40 % y el 25 %. De los envases de vidrio que llegan para reciclar de A, el 70 % son botellas, de los de B, lo son el 65 %, y de los del C, el 80 %. A partir de esta información se pide:

    (a) Si cogemos una botella de vidrio al azar de las llegadas a la planta de reciclaje, ¿de qué punto de recogida es más probable que haya llegado?

    (b) Si cogemos al azar un objeto de vidrio de los llegados a la planta de reciclaje, ¿cuál es la probabilidad de que sea *"una botella y haya llegado del punto de recogida C"*?

    (c) Si cogemos ahora al azar 10 objetos de vidrio de los llegados a la planta de reciclaje, ¿cuál es la probabilidad de que de los 10 haya como máximo 4 que sean *"botellas y hayan llegado del punto de recogida C"*, si sabemos que al menos 3 de ellos lo cumplen?

11. Los datos de la tabla muestran la *Magnitud* (en la escala Richter) y la *Profundidad del foco* (en Km.) para un total de 9 terremotos registrados en cierta zona geográfica durante un mes:

Magnitud	1.95	1.82	1.65	1.88	2.05	1.95	1.93	1.90	1.89
Profundidad	5	7	9	7	5	6	6	6	7

Se pide:

(a) Calcular el coeficiente de correlación (lineal) de Pearson entre las dos variables, *Magnitud* y *Profundidad del foco*. ¿Podemos decir, con una probabilidad de equivocarnos de, como mucho, 0.05 al hacerlo, que hay correlación lineal negativa significativa entre las dos variables?

(b) Obtener la recta de regresión adecuada para predecir la Profundidad del foco de un terremoto que se ha producido en la zona y cuya Magnitud en la escala de Richter ha sido de 1.75, si es que es posible realizar la predicción.

12. Del total de alumnos matriculados en la Enseñanza Universitaria en Cataluña durante el curso académico 2005/06, el 57 % eran mujeres. Se sabe que sólo el 11 % de las mujeres que estudian en la universidad cursan titulaciones técnicas, mientras que este porcentaje sube hasta el 31 % para los hombres. Estos datos son reales, aunque redondeados, y han sido obtenidos del *Anuari Estadístic de Catalunya 2007* publicado por el *Institut d'Estadística de Catalunya, Generalitat de Catalunya.* A partir de estos datos se pide:

(a) De los alumnos matriculados en carreras técnicas el curso 2005/06 en Cataluña, ¿cuál es el porcentaje de mujeres?

(b) De los alumnos universitarios en Cataluña durante el curso 2005/06, ¿qué porcentaje son *mujeres o cursan una titulación técnica*?

(c) De los alumnos universitarios en Cataluña durante el curso 2005/06, ¿qué porcentaje son *hombres y no cursan una titulación técnica*?

13. Se sabe que el *valor predictivo positivo* (probabilidad de que realmente tenga la enfermedad un individuo que da positivo en la prueba) para una prueba clínica depende de la *prevalencia* en la población de la enfermedad que la prueba pretende detectar, digamos $p \in (0,\ 1)$.

    Si la prueba clínica tiene una probabilidad de "falso positivo" α, y una probabilidad de "falso negativo" β $(0 < \alpha, \beta < 1)$, y se desea que el *valor predictivo positivo* no sea inferior a un valor de referencia que se considera mínimo para que sea aceptable usar la prueba clínica en esa población, $v \in (0,\ 1)$, ¿cómo debería ser p para que esto se cumpliese?

    ¿Es o no más fácil que se cumpla esto para las enfermedades llamadas "raras" que para las otras?

    Aplicación: Encontrar cómo debe ser p si $\alpha = 0.01$, $\beta = 0.05$ y $v = 0.95$.